Journal of the Association for Information Systems

JAIS

Research Article

A New Approach to Testing Nomological Validity and Its Application to a Second-Order Measurement Model of Trust

Liping Liu
The University of Akron
liu@acm.org

Chan Li
The University of Pittsburgh
chanli@katz.pitt.edu

Dan Zhu
Iowa State University
dzhu@iastate.edu

Abstract

This paper examines the concept of the nomological validity of second- and/or higher-order measurement models. It also proposes a new approach that consists of measuring two validity indices — predictive and mediating efficiencies — to compare the efficacies of a research model with and without a higher-order abstraction. To illustrate this concept, we test a second-order measurement model of trust and study how it behaves in a nomological network of the consumer's prior experience as an antecedent and willingness to buy as a consequent variable.

Keywords: *Nomological Validity, Ecommerce, Trust in Internet Stores, Higher-Order Factor Models.*

* Carol Saunders was the accepting senior editor. This article was submitted on 5th March 2010 and went through four revisions.

Volume 13, Issue 12, pp. 950-975, December 2012

A New Approach to Testing Nomological Validity and Its Application to a Second-Order Measurement Model of Trust

1. Introduction

In information systems (IS), many phenomena are complex and multidimensional in nature and cannot be properly measured by a simple first-order latent construct (DeLone & McLean, 1992). Therefore, higher-order models have been proposed to capture these phenomena. Examples include end-user satisfaction (Doll & Torkzadeh, 1991), information system planning success (Segars & Grover, 1998), service quality (Jiang, Klein, & Carr, 2002), information privacy (Stewart & Segars, 2002), trust (Benbasat & Wang, 2005; McKnight, Choudhury, & Kacmar, 2002), and perceived justice for information provision (Son & Kim, 2008). Some of these higher-order models have been questioned for their construct validity. A particular issue of interest is concerning the milieu in which a model may be validated. A higher-order model cannot exist in insularity. It needs to relate to other factors or be placed in a nomological network of consequent and/or antecedent variables to determine if it acts as a better mediator than its underlying first-order factors. Such an aspect of measurement efficacy is called "nomological validity".

Nomological validity is a relatively new concept in organizational research in general and IS in particular; techniques to conduct the tests or criteria to make judgments have not, as yet, been well developed. In the IS literature, Chin (1998) has proposed embedding a second-order latent variable as a mediator between an antecedent variable and a consequent variable; unfortunately, Chin describes neitherthe process nor the criteria. Other studies have discussed nomological validity (McKnight et al., 2002; Salisbury, Chin, Gopal, & Newsted, 2002; Stewart & Segars, 2002), but no study has provided a formal treatment of the concept. Many studies have used the term for a different concept (e.g., Kappelman, 1995). Among researchers who have conducted tests, the procedures and criteria have been implicit and inconsistent. There have also been differences even in terms of how nomological validity is defined.

In this paper, we systematically examine the nature of nomological validity and analyze the difficulties associated with the test procedures used in existing IS studies. We then propose a new approach based on the explanatory power of a measurement model. In particular, we propose the notion of "predictive efficiency" to measure how well higher-order factors as a predictor explain the variance in a consequent variable compared to first-order factors as joint predictors. We propose the notion of "mediating efficiency" to measure how well an antecedent explains the total variance in a concept indirectly through the mediating effect of a higher-order factor compared to when it does so directly without the mediating variable. Both efficacy indices should be above certain thresholds to support the nomological validity of a higher-order factor model.

We further apply our method to the test of the nomological validity of a second-order measurement model of trust, which we empirically developed from a controlled experiment involving 173 subjects and three online bookstores. In this model, we conceptualize trust in Internet stores as a second- order factor manifested by four first-order latent factors: inopportunism, privacy, security, and transactional accuracy. We consider a consumer's prior experience as an antecedent to trust and their willingness to buy from an Internet store as a consequent variable. We empirically test the model using usual test procedures with an emphasis on the illustration of our new approach to nomological validity.

The rest of the paper is organized as follows. Section 2 provides a conceptual background on the nomological validity. Section 3 proposes our new approach to testing nomological validity. Section 4 describes the application of our approach to a measurement model of trust. Section 5 concludes our paper with highlights of our contributions and points out the limitations and directions for future research.

2. Background

Development of a measurement model must be preceded by a sound conceptual specification of the construct to be measured (Churchill, 1979; Netemeyer, Bearden, & Sharma, 2003). The key is a "well-articulated" theoretical foundation that justifies the relationships between the construct and its underlying dimensions (Chin, 1998). According to how such relationships are conceptualized, a measurement model can be generally placed into two broad categories (MacKenzie, Podsakoff, & Jarvis, 2005): formative or reflective.

In a formative model, each dimension taps a unique aspect of and measures a unique portion of the overall construct to be measured, and all the dimensions combine to produce the aggregate construct. In research models, each dimension is a causal node pointing to the aggregate construct (see A in Figure 1 for the formative first-order model). In contrast, in a reflective framework, each dimension is considered to be a manifestation of the construct to be measured (DeVellis, 2003). All dimensions are highly correlated, which reflects the same construct to be measured (Bollen & Lennox, 1991), and each dimension samples the “whole” concept or “superordinate construct” (Edwards, 2001; Edwards & Bagozzi, 2000). In research models, each dimension is shown as an effect node emanating from the superordinate construct (see B in Figure 1).

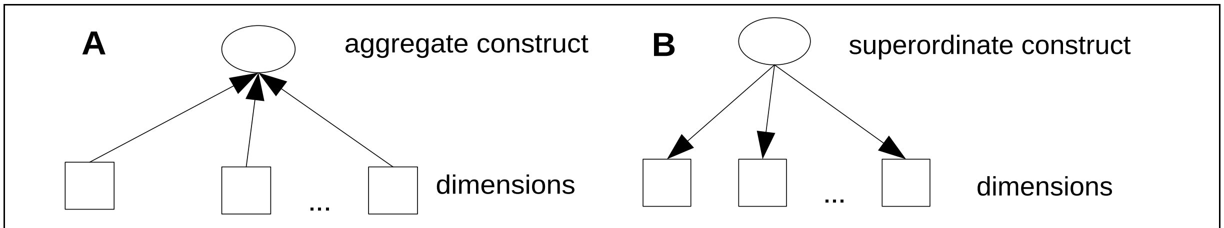

Figure 1. One-Factor Formative and Reflective Models

The formative and reflective structures also apply to second- and higher-order constructs. For example, Model A in Figure 2 shows a formative structure between an aggregate second-order construct and its first-order components or dimensions. Such formative structures are seen in studies on organizational trust (Mayer, Davis, & Schoorman, 1995), data quality (Wang & Strong, 1996), consumer trust (Lee & Turban, 2001), attitude towards technology innovation (Taylor & Todd, 1995), and perceived consequence of behavior (Thompson, Higgins, & Howell, 1991). Research models often place an aggregate construct as a set of first-order factors in the form of independent components (see Model B in Figure 2), and hypothesize its relationship with an antecedent or consequent variable by stipulating multiple relationships between each first-order component and the antecedent/consequent. Sometimes, not all components are stipulated to have the same antecedents or consequents (MacKenzie et al., 2005). Therefore, in a formative structure, the measurement of a concept as one higher-order construct appears unnecessary.

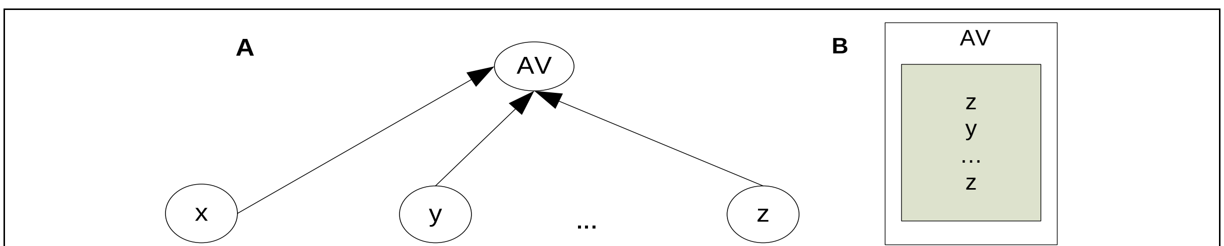

Figure 2. A Higher-Order Formative Factor Model

Since there can be either a formative or a reflective structure for either a first-order or a higher-order factor, there is a continuum of structures as measurement models (MacKenzie et al., 2005). Among them, the standard higher-order factor model is made of a higher-order superordinate reflected by lower-order dimensions, and each dimension is in turn reflected by measureable indicators (Rindskopf & Rose, 1988). Here, the dimensional factors may refer to narrowly defined phenomena or fine-grained aspects of some broader construct, whereas the higher-order factor is meant to capture a global holistic phenomena (Bagozzi & Edwards, 1998).

The nomological validity studied here pertains to higher-order reflective models. Because the higher-order factor is the only common covariate of the dimensional factors, it is supposed to capture the covariance among all first-order factors. As such, the higher-order reflective model has two advantages (Bagozzi & Edwards, 1998). First, it is able to capture the underlying meaning of a concept holistically rather than that of its piecemeal dimensions. Second, it represents the overall concept more parsimoniously when compared to first-order factors; the overall construct appears in a research model as a single latent variable rather than a set of variables. However, the advantages come at a price, and it is likely that a reliable scale can be achieved but be invalid. For example, the construct as measured may be an artificial entity rather than the intended concept (Bagozzi & Edwards, 1998). Therefore, it is imperative to link a higher-order model to related concepts and check against such possibilities.

The key to a nomological test is to embed a high-order factor model into an extended nomological network that comprises antecedent and/or consequent variables and see how well it behaves relative to its underlying first-order factors. The judgment about nomological validity involves four models (see Figure 3).

- Model A is the group-factor model (Rindskopf & Rose, 1988). It assumes that a group of first-order factors represent a complex concept and account for the covariance among all indicators. The first-order factors may correlate with each other but the correlations are not due to unspecified common causes.

- Model B is the standard second-order model (Rindskopf & Rose, 1988). It assumes that first-order factors account for the variance in the indicators but that a second-order factor accounts for the variance in the first-order factors. The first-order factors correlate due to a common cause – the second-order factor that represents the concept to be measured.

- Model C extends model A by connecting the group-factor model directly to antecedent and consequent constructs. As Figure 3 illustrates, the antecedent and consequent variables are shown in gray circles with measurable indicators and inserted at the top of the group-factor model. Like the group-factor model, model C assumes that the concept is represented by a group of first-order factors, which correlate due to unspecified common causes. It also hypothesizes each first-order factor as a mediator between the antecedent and the consequent, which the bold links between each gray circle and each of the first-order factors in the group-factor model show.

- Model D extends model B by connecting the standard second-order factor model to antecedent and consequent variables (shown in gray circles). Unlike model C, here we hypothesize the second-order construct rather than individual first-order factors as a mediator between the antecedent and the consequent, which the bold links between the second-order construct with each of the gray circles show.

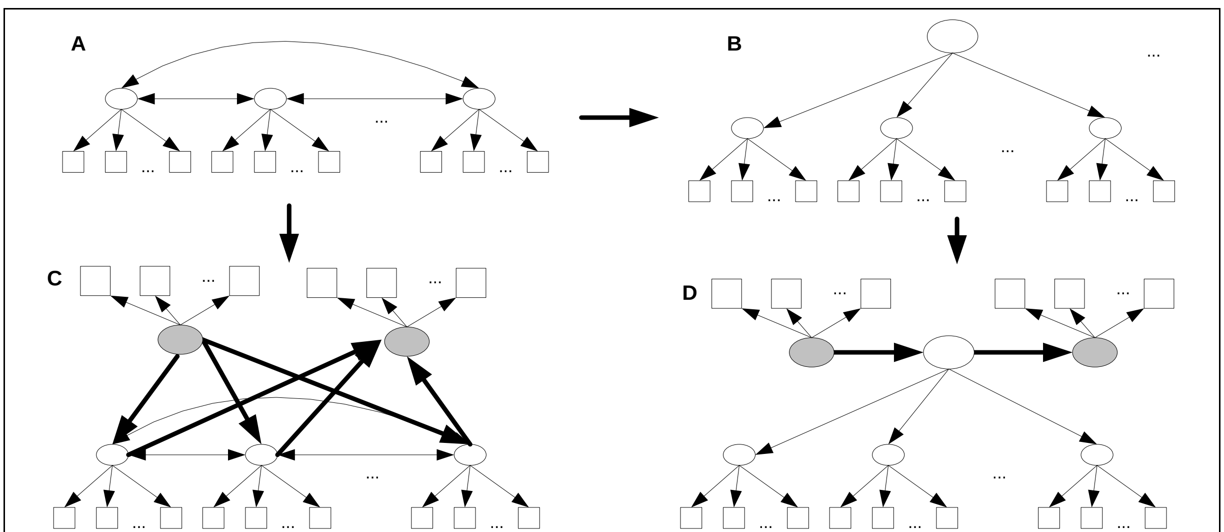

Figure 3. Alternative Models for Nomological Validity Testing

Model B is the target whose nomological validity is to be assessed, whereas Model A acts as a benchmark for the assessment. Although a higher-order model is able to explain the covariance of first-order factors, the goodness-of-fit of the higher-order model can never be better than that of the corresponding group-factor model (Marsh & Hocevar, 1985). Therefore, model A is often chosen as the benchmark to be compared with model B. In structural equation modeling, the one-factor model (see B in Figure 1) is often used as the benchmark against group-factor or higher-order factor models (Rindskopf & Rose, 1988). Prior to assessing the nomological validity of a higher-order factor, it is assumed the higher-order factor model is already identified and comprises multiple dimensions. Therefore, a multifactor or group-factor model should be the benchmark. In addition, without loss of generality, one may treat the one-factor model as a special case of the group-factor model; the concept of and the approach to nomological validity will be the same.

Existing studies have used many different criteria to test nomological validity. At least three approaches appeared in the IS literature. Steward and Segars (2002) compared model D with their benchmark model, which is similar to model C with the exception of having no correlations among the first order factors. They suggest that an equal or better fit of model D relative to that of the benchmark model supported the nomological validity of model B1. We call such a criterion the "relative model-fit" one. McKnight et al. (2002) did not use a benchmark; they suggest that a construct exhibits adequate nomological validity if it is strongly correlated to an antecedent or consequent construct. We call such a criterion the "absolute model-fit" one. Salisbury et al. (2002) tested the nomological validity of a first-order construct. Their approach was to compare model C with model A to see whether there are big changes in model fit indices such as item loadings, reliabilities, and convergent and discriminate validities. Little or no change supports nomological validity. Of course, one may apply the same approach to the second-order construct and compare model D to model B. We call such a criterion the "model-fix" one.

The existing approaches make intuitive sense; for example, model D has a simpler structure than Model C and is thus a more viable model according to Occam's razor if it fits the data equally well or better than Model C. However, due to its ad hoc nature, the concept of nomological validity has never been consistent and is sometimes difficult to be generalized.

First, there are exceptions to the relative model-fit criterion. In theory, a simple model, while not as accurate as a complex one, can still be accepted if it fits data well. It is well known that, in linear regression, adding more predictors will lead to increased R2 values and "improved" explanation power. However, one does not simply go with a model by maximizing its R2 value. A structural equation model essentially consists of multiple simultaneous regression models. Therefore, its model selection criterion also entails a tradeoff for structural parsimony.

In practice, there are cases in which model D has better model-fit than model C and vice versa. Therefore, the relative model-fit may not indicate nomological validity properly. One case is when there are no consequent variables, all first-order factors highly correlate with each other, and the second-order construct captures a concept better than any of its first-order factors. In this case, the antecedent predicts the second-order factor better than any first-order factor, and the second-order construct better predicts first-order variables than the antecedent.Thus, model D has a better model-fit than model C due to its increased prediction powers. The opposite case is when there is no antecedent variable or when there are many first-order factors. If no antecedent exists, model C is more likely to fit data better than model D because the former uses all first-order factors as joint predictors whereas the latter uses the second-order construct – which captures only a portion of the variance in the first-order factors. If there are many first-order factors, there will likely be more parameters[2] estimated in model C than in model D; therefore, one may expect increased model-fit indices for model C relative to model D.

[1] If their base model had included the correlations as model C does, their result may not have met this criterion because their model C will have reduced fit indices.

[2] The increase is $n(n-1)/2 + n - 4$, where n is the number of the first-order factors.

Second, it is technically difficult to ensure the invariance of a measurement model to meet the model-fix criterion. In theory, a construct represents a concept with definite semantics and attributes. It must maintain the same measure regardless of whether it stands alone or is embedded into a nomological network of other variables. This entails little or no change of factor loadings in confirmatory factor analysis (i.e., each path coefficient from the construct in question to a dimensional variable must be invariant with or without the presence of external variables). Unfortunately, this model-fix criterion cannot be upheld when a higher-order construct is embedded into a nomological network of consequent variables. Technically, a consequent is a reflective factor of the construct to be measured and is equivalent to a dimensional variable in causality. Thus, in confirmatory factor analysis, connecting a higher-order construct to consequent variables redefines the higher-order construct and expands the original measurement model by practically transforming each consequent variable into a new dimension. As such, the construct, after being connected to its consequents, carries neither the same meaning nor the same measure as when it stands alone. Significant changes in factor loadings will be inevitable as evidenced in existing studies (e.g., Stewart & Segars, 2002). Our simulations show that the magnitude in loading change depends on the correlations between the construct in question and the consequent variables; the smaller the correlations, the greater the change.

Third, the absolute model-fit criterion has conceptual difficulties. The higher-order is meant to capture the essence of the underlying meaning of its dimensions. However, as Bagozzi and Edwards (1998) point out, the higher-order factor is likely conceptually invalid; it may be an artificial entity rather than the intended concept as reflected by its dimensions. A large correlation guarantees only that the higher-order construct is related to other variables but does not guarantee that it is not an artificial entity. Therefore, it is imperative that we embed antecedent and consequence variables into not only the higher-order factor but also the group-factor model and compare how well the higher-order model performs relative to the group-factor model.

In summary, the model-fix criterion focuses on the stability of a measurement model; however, it is not applicable to a higher-order factor because, when a variable is modeled as a causal consequent of a higher-order factor, it automatically becomes another dimension of the higher-order factor and the model change is inevitable. The two model-fit criteria test whether a higher-order model fits data well after they are embedded into a larger network; however, it does not indicate whether the model is still the same after being related to other variables. In addition, the absolute model-fit criterion does not exclude the possibility that the higher-order construct may be an artificial entity rather than the intended concept.

3. A New Approach

To overcome difficulties in existing approaches, here we propose a new approach to nomological validity. Our test comprises a two-part power analysis. The first part computes the power of a higher-order variable relative to its first-order factors in predicting a consequent variable. The result is called "predictive efficiency". The second part computes the power of an antecedent in predicting the first-order factors indirectly through the second-order variable as a mediator compared to when it does so directly without the mediator, which results in "mediating efficiency". Both efficiencies have an upper limit of 100 percent, with higher values supporting nomological validity.

As with other approaches, the first step is carried out to ensure that antecedent and consequent variables are properly sampled from the context, and each factor in the measurement model is properly measured. The usual test procedures and criteria apply here, which include content, convergent, and discriminant validity. Bohrnstedt (1970) defines content validity as the degree to which the scale being used represents the concept about which generalizations are to be made. Psychometrics often appeals to the so-called domain sampling technique to ensure content validity (Nunnally, 1978). Because of the reflective structure, if multiple scale items measure the same construct, there should be convergent validity – items meant to measure the same construct are highly correlated. The first-order dimensions are also highly correlated, which reflects the same higher-order construct to be measured (Bollen & Lennox, 1991), and each dimension samples the "whole" concept (Edwards, 2001; Edwards & Bagozzi, 2000). A popular measure of convergent validity is Cronbach's alpha coefficients. On the other hand, items that measure different constructs must possess discriminant validity – an item should correlate more highly with other items meant to measure the same construct than with items used to measure a different construct (Campbell & Fiske, 1959). Discriminant validity may be tested using principal components analysis with orthogonal or oblique rotations. It should be applied to all of the first-order constructs that reflect the higher-order construct along with the antecedent and consequent variables to be included in the nomological validity test.

The second step is to freeze the measure of the construct to ensure that it has identical semantics and properties before and after it is embedded into a nomological network of other variables. A measure can be frozen by using factor scores as the proxy observations of the construct. The factor score estimates can then be computed using either exact or coarse methods. Exact methods yield approximately standardized factor score estimates that achieve determinacy (Thurstone, 1935) or correlation-preserving (Anderson & Rubin, 1956). These methods are implemented in statistical packages, including SPSS and SAS. Coarse methods compute estimates by simply summing the responses of subsets of the factored items (Cattell, 1952). For example, the score of a first-order factor is the average of the responses to the items reflecting the factor. The score of a second-order factor is the average factor score of those reflecting the second-order factor.

3.1. Predictive Efficiency

The objective of using a higher-order factor is to represent a complex concept parsimoniously; that is, using a single construct rather than multiple first-order constructs. Consequently, an important question concerning nomological validity is whether the higher-order construct can sufficiently predict a consequent variable, which supposedly can be predicted jointly by the underlying first-order factors. To answer this question, we ran two linear regressions as diagrammed in Figure 4:

(1) The regression of the consequent variable against all first-order variables, and
(2) The regression of the consequent against the higher-order construct.

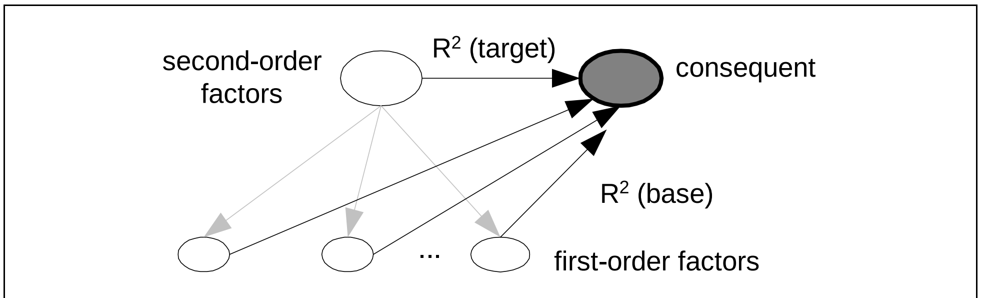

Figure 4. Computing Predictive Efficiency

Let R^2 (base) denote the amount of variance in the consequent variable explained by the first-order variables jointly. Let R^2 (target) denote the variance explained by the higher-order factor solely. The ratio of R^2 (target) to R^2 (base) then measures how well the higher-order variable relative to the first-order variables explains the variance in the consequent variable. We call such a ratio "predictive efficiency". Intuitively, a higher-order factor captures the common portion of the variance in the first-order variables. Thus, among all the variances in the consequent variable that can be jointly explained by the first-order factors, only a portion may be explained by the higher-order factor. Thus, R^2 (target) is less than or equal to R^2 (base). The higher the predictive efficiency is, the more variance is explained, and therefore the better higher-order factor replaces the role of the first-order factors in predicting a consequent variable. To summarize these results, we state the first nomological validity criterion as follows:

Criterion 1: *The predictive efficiency of any higher-order construct is less than or equal to 100 percent with higher values indicating better nomological validity.*

3.2. Mediating Efficiency

Assume that the domain of a multi-dimensional concept is fully covered by its first-order factors. Then the total variance inherent in the concept is the sum of the variance inherent in these factors. Without a higher-order factor acting as the proxy of the concept, we may use an antecedent to predict the concept directly by regressing each of the first-order factors against the antecedent. However, with a higher-order factor representing the concept, we use the antecedent to predict the concept indirectly first through the regression of the second-order factor against the antecedent and then through the regression of the first-order factors against the second-order factor (see Figure 5). The second part of the nomological test considers how well this indirect prediction, relative to the direct prediction, explains total variance in the concept.

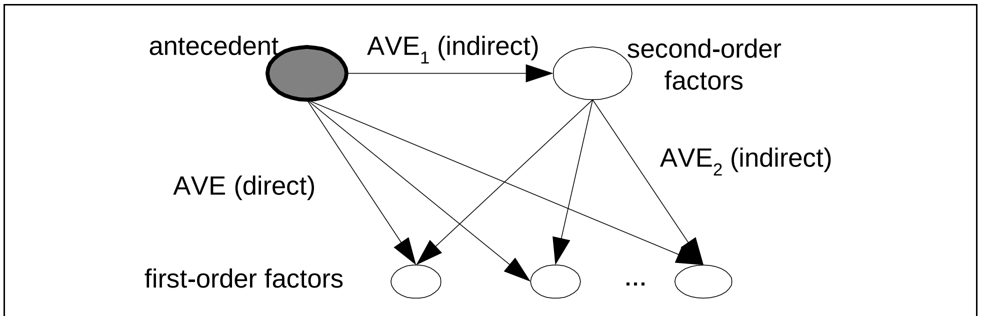

Figure 5. Computing Mediating Efficiency

For direct prediction, we run multiple regressions, one on each first-order variable against the same antecedent. We use the concept of average variance extracted (AVE) proposed by Fornell and Larcker (1981) to account for the total variance explained by the regressions. Note that, when multiple regressions predict different dependent variables using the same predictor, the AVE is not, as its name suggests, a simple average of the variances explained by the individual regression models. Instead, it is the ratio of total explained variance to the total variance including both explained and unexplained variances in the dependent variables[3]. In other words, among the total variance in all the dependent variables, AVE represents the portion that can be explained by the predictor. Applying the concept to the direct prediction of a concept, the corresponding AVE, denoted by AVE (direct), represents the amount of variance in the concept that is explained by the antecedent.

For indirect prediction, the antecedent predicts the higher-order variable and explains a portion of the total variance in the variable. Let AVE_1 (indirect) denote such a portion of the variance. AVE_1 (indirect) is simply the R^2 value of the regression on the higher-order variable against the antecedent. Furthermore, due to its reflective structure, the higher-order factor model has already assumed that the higher-order factor predicts each of its manifested first-order factors. The corresponding AVE represents the reliability of the higher-order factor in the context of testing measurement validity. Now the same number represents the amount of the total variance inherent in the concept that can be explained by the higher-order factor. Let AVE_2 (indirect) denote this portion of the variance. By combining both steps, the antecedent explains the AVE_1 (indirect) percent of the variance in the higher-order variable, which, in turn, predicts the AVE_2 (indirect) percent of the variance in the concept to be measured. Their product, AVE_1 (indirect) $\times$ AVE_2 (indirect), is then the percentage of the total variance in the concept that is indirectly explained by the antecedent.

[3] Assume λ_i and $Var(\varepsilon_i)$ are respectively the regression coefficient and error variance for i^{th} regression, then the AVE of all the regressions is $\sum \lambda_i^2 \Big/ \sum(\lambda_i^2 + \mathrm{var}(\varepsilon_i))$.

Journal of the Association for Information Systems Vol. 13, Issue 12, pp. 950-975, December 2012

Therefore, among the total variance in a concept, AVE (direct) represents the portion of what an antecedent can explain directly. AVE_1 (indirect) × AVE_2 (indirect) represents the portion of what the same antecedent can explain indirectly. The ratio of AVE_1 (indirect) × AVE_2 (indirect) to AVE (direct) then measures how well the indirect prediction relative to the direct prediction explains the variance in the concept. This ratio is called “mediating efficiency”.

In terms of causality, the higher-order factor plays the role of a mediator between an antecedent and each first-order variable. The path coefficient between the antecedent and a first-order variable is greater than the products of the path coefficient between the same antecedent and the higher-order variable, and the path coefficient between the higher-order variable and the first-order variable. Thus, AVE_1 (indirect) × AVE_2 (indirect) ≤ AVE (direct). Of course, the closer the two sides are, the better the higher-order factor replaces the role of the first-order factors in terms of how the underlying concept is predicted by a causal antecedent. Therefore, the second nomological criterion is:

> ***Criterion 2:*** *The mediating efficiency of any higher-order construct is less than or equal to 100 percent, with higher values supporting nomological validity.*

To recap, for a measurement model to be nomologically valid, both predictive and mediating efficiencies must have high values. Any index with a low value will indicate a possibility that the higher-order factor is an artificial entity or does not represent the concept as defined by the first-order dimensions. When this happens, a researcher may have to abandon the higher-order construct; they could either use the group-factor model or redefine the higher-order construct using a different set of dimensions.

The lower the efficacies, the higher the possibility of getting an artificial entity. Determining what is low, high, or acceptable, or whether the threshold should be set at 95 percent, 90 percent, or 70 percent, is a judgmental call. It is like the situation for determining the threshold for marking stars in hypothesis testing; should the significance level be 0.01, 0.05, or 0.1? Note that both predictive and mediating efficiencies are defined as percentages of variance retained; their thresholds cannot go below 50 percent or the higher-order construct loses more explained or captured variance than it retains. The more reasonable threshold may be above 75 percent, which means that the higher-order construct loses no more than a quarter of the variance explained or captured. Again, this is a tentative suggestion.

4. The Nomological Validity of Trust

The concept of mediating and predictive efficacies applies to reflective higher-order measurement models. Many existing models in the literature lack clear distinctions between formative and reflective structures (MacKenzie et al., 2005). To properly test our new method, we developed a second-order measurement model of trust from the group up and empirically validated it from a controlled experiment involving 173 subjects and three online bookstores

4.1. A Second-Order Measurement of Trust

Trust is a psychological state comprising the intention to accept vulnerability based on positive expectations of the behavior of another (Rousseau, Sitkin, Burt, & Camerer, 1998). Risk is a necessary condition for trust to arise (Lewis & Weigert, 1985). In Internet shopping, trust is often an order qualifier for purchase decisions (Doney & Cannon, 1997). The vulnerability is reflected in four specific dimensions, which includes security, privacy, merchant opportunism, and transaction accuracy. Trust or a lack of it depends on the ability and the responsibility of the Internet store in managing these risks (Barber, 1983). Therefore, trust in an Internet store is defined as an individual’s beliefs or perceptions of security, privacy, opportunism, and transaction accuracy that reflect the merchant’s ability and responsibility in managing the concerns. The perceptions are conceptualized as “perceived security” (PS), “perceived privacy” (PP), “perceived inopportunism” (PI), and “perceived accuracy” (PA), respectively.

We should note that this proposed definition of trust differs from that used in some existing studies (Benbasat & Wang, 2005; McKnight & Chervany, 2002), which define trust as manifested beliefs on a merchant’s competence, benevolence, and integrity. However, there are two issues with these studies. First, it is arguable whether dimensions such as benevolence are characteristics of online merchants or agents; merchants are not charities, and consumers do not expect them to be benevolent to gain trust. What is of concern to consumers (or what matters to trust) is whether the merchant follows a set of fair procedures and principles in conducting businesses. Second, a merchant can be competent but may lack benevolence or integrity, or may be benevolent but incompetent; these dimensions do not necessarily correlate. Thus, they do not form the standard reflective structure of higher-order models according to the criteria of MacKenzie et al. (2005). In fact, Mayer et al. (1995) originally used competence, benevolence, and integrity as formative, rather than reflective, components of organizational trust. The above studies on trust (Benbasat & Wang, 2005; McKnight & Chervany, 2002), however, adopted the concept from Mayer et al. (1995) but turned around the causality between trust and its dimensions and made the same dimensions reflective.

In contrast, our proposed definition equates trust to a merchant's inherent qualities such as ability and responsibility. In the parlance of the integrative model of Mayer et al. (1995), trust comprises ability and responsibility. If the store is capable and responsible (i.e., trustable), the buyer's concerns in all the four concerned area will be low. On the other hand, if it is not capable and responsible, the buyer's concerns will increase, which reflects a lack of trust. Therefore, perceived security, privacy, inopportunism, and accuracy are reflective dimensions of trust (see Figure 6).

4.1.1. Perceived Security

In the context of e-commerce, communication can be intercepted, tampered with, and falsified. Personal accounts can also be accessed for illegal purposes. Concern with security has become the top reason for not shopping online (Hoffman, Novak, & Peralta, 1999). To reflect the intention to accept the vulnerability in security, perceived security is conceptualized as the perception that making a transaction with an Internet store is safe. Such a notion of security signals the capability and responsibility of an Internet store to prevent unauthorized data access, illegal data interception, and illegal attacks on customer properties, and protect the interests of its customers. Of course, absolute security does not exist. Thus, perceived security is operationalized in terms of a comparison with traditional means of shopping and whether transmitted data can be intercepted and customer accounts compromised.

4.1.2. Perceived Privacy

Privacy usually refers to personal information and its confidentiality. For example, when customers give out their identification numbers, they expect confidentiality; that is, that the numbers will be known only by the party who has a legitimate need to know them. Similarly, a customer who buys a bulk of marketing research data may not want their competitors to know it, and a customer who buys a good of questionable value does not want to disclose their identity. The violation of privacy includes unauthorized collection, disclosure, or other misuse of personal information as a direct result of e-commerce transactions. Privacy has been identified as one of the most crucial issues in e-commerce. Consumers' fear and distrust for potential loss of personal privacy often makes them unwilling to conduct online transactions (Son & Kim, 2008; Wang, Lee, & Wang, 1998). To reflect their intention to accept the vulnerability in privacy, we conceptualize perceived privacy as the consumer's perception that their personal data with an Internet store will be kept confidential. The concept signals the degree to which an Internet store is capable of and responsible for observing procedural fairness (Hoffman et al., 1999) and to exert control on its data collection, access, and secondary use (Stewart & Segars, 2002). It represents a fundamental concern for the loss of proper control on personal data due to improper collection of private information, improper monitoring on Internet activities, and improper transfer of personal information to other businesses.

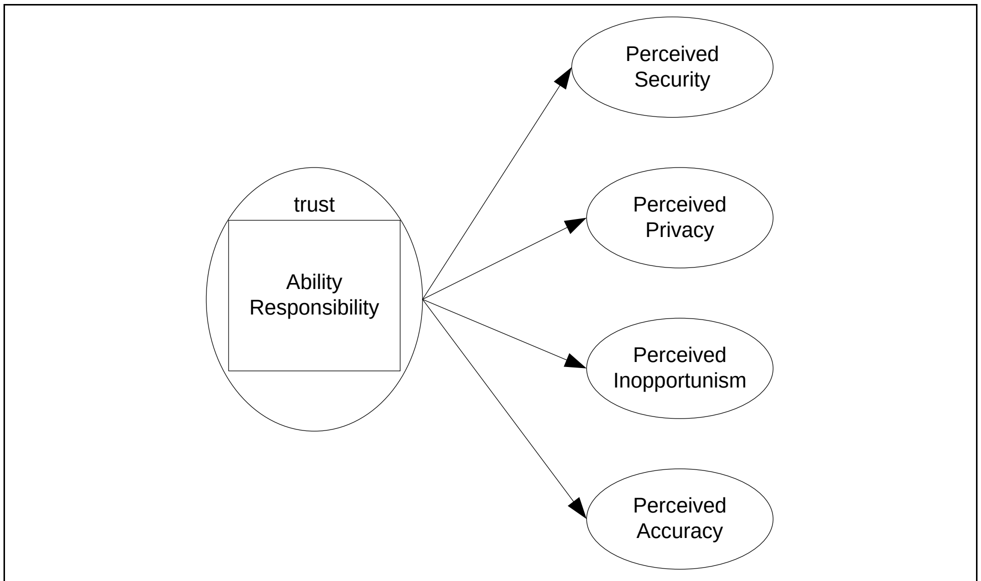

Figure 6. Trust and its Dimensions

4.1.3. Perceived Inopportunism

When making online transactions, customers deal with other parties whose true motives and absolute identity are uncertain (Tygar, 1998). They are concerned about whether the store will take advantage of them and behave opportunistically. For example, they are concerned about not receiving products or services although they have paid for them. They are also concerned about whether the store will deliver the products and services as described or substitute fake or inferior products instead. To reflect such concerns and the intention to accept the vulnerability in opportunism, perceived inopportunism is defined as one's belief that an Internet store is honest in adhering to an acceptable set of principles. This definition involves two related perceptions: (a) perceptions of being procedurally fair (i.e., that a store has a fair procedure and follows it), and (b) perceptions that the store is honest and credible (i.e., that the store keeps promises and acts and says consistently).

Note that this concept of inopportunism is sometimes defined or treated identically in the literature as integrity (Lewis & Weigert, 1985; Smith & Barclay, 1997). However, integrity generally has a connotation of deceptions rather than fair procedures. In addition, the concept has been established as a formative factor of (or antecedent to) trust (Mayer et al., 1995). To avoid possible confusion, we coin the word inopportunism as "forbearance from opportunism".

Note also that this conception parallels similar discussions of reputation that emphasize behaving in a fair, honest, and consistent manner. For example, Doney and Cannon (1997) define reputation as the extent to which customers in the industry believe that a company is honest and concerned about its customers. Since reputation has a connotation of public opinions or evaluations of a store, it is inappropriate concept to represent a customer's beliefs on opportunism.

4.1.4. Perceived Accuracy

Besides opportunistic behaviors, customers are also concerned about potential transaction errors such as incorrect product brands, sizes, and quantities, as well as incorrect billing statements. Such errors occur due to human or mechanistic mistakes than opportunism. They include computer system irregularities such as lack of transaction atomicity (Tygar, 1998). In physical commerce, such errors can be corrected with a trip or phone call back to the store. However, making corrections with an online store often means extra effort and time and sometimes even extra shipping and handling fees. Therefore, Jarvenpaa, Tractinsky, and Vitale (2000) suggest that, in order to be able to trust an Internet store, a customer must believe that the store has both the ability and the motivation to reliably deliver goods and services of the quality expected. Similarly, Butler (1991) introduces the dimension of transaction accuracy by emphasizing the criteria of trust such as consistency and discreetness. To capture such a dimension of trust, perceived accuracy is conceptualized as the extent to which a customer believes that transactions with an Internet store are error-free. This concept signals the capability and responsibility of an Internet store in performing business functions accurately. Note that the notion of accuracy is related to but narrower than service reliability. We believe it entails two or more shopping instances for reliability measures to be meaningfully rated by the participants. Considering our experimental setting, we settled with the construct of perceived accuracy.

In sum, a customer's trust in an Internet store is reflected by their beliefs about whether their transaction will be carried out securely, confidentially, non-opportunistically, and accurately. To operationalize these beliefs, each unique and meaningful concern is casted in the form of a single sentence that reflects the corresponding theoretical dimension of trust. Following the advice by Churchill (1979), we utilized the expert panel and a class of undergraduate students to participate in a pre-test and a pilot test respectively. In the pre-test, we provided a formal definition of each construct and a list of measurement items (sentences), which we intended to use to measure the construct. We asked each member to first read each definition carefully and then give a rating for each item on a five-point scale to indicate how well the sentence matched the intended construct. In addition, we asked them to rewrite the sentence in a clearer way or make a suggestion right on or next to an item if they felt it was awkward and ambiguous.

Because our item development at this stage was exploratory in nature and none of the four dimensional constructs had been developed in the literature, we employed the semantics-matching technique to confirm existing items or generate better ones. The pre-test started with 48 items in total. After the pre-test, we analyzed the ratings of each item individually. An item was retained if it was consistently rated four or five by all of the experts. An item was dropped if it was consistently rated as poorly matched with the construct. For an item that had inconsistent ratings, we either adopted an alternative provided by one of the experts, rephrased it, or dropped it. The pretest substantially refined some of the items by eliminating their ambiguity. The total number of items was also reduced to 33.

To further validate the items, we conducted a pilot test using 35 undergraduate students. We randomized the 33 items and created a survey that asked each participant to visit an online store and respond to each item by indicating how much they agreed with the item's statement. Then we conducted a content analysis of the items. Under each dimensional construct, we retained items that were highly correlated for internal consistency. Depending on their content, we either dropped or modified items that were not highly correlated,. As a result of the pilot test, we selected 17 items for inclusion in the final instrument (see Appendix A-1).

The final test involved 173 participants selected from graduate and undergraduate students in two large national universities. In survey research, a damaging confound is a lack of knowledge by respondents (Segars & Grover, 1998). To reduce such a confounding effect, all statements were rephrased using less technical jargon, and terms like encryption and integrity were avoided. Nevertheless, if a participant was not aware of Internet technology and its potential problems, they may not have made perfect sense of some of the statements in Appendix A-1. Therefore, when identifying participants, we required them to have exposure to e-commerce. To be representative of the population actually engaging in Internet commerce activities, we identified participants from students at various educational stages: 24 percent from graduate programs, 38 percent from juniors and seniors, and 38 percent from freshmen and sophomores. The percentages were consistent with the distribution of online student shoppers according to the Internet usage database maintained by Georgia Tech. We also selected participants equally from both rural and urban areas. We primarily selected participants from four graduate and eight undergraduate classes to participate on a voluntary basis. Among those individuals qualified to participate, the response rate was 73 percent.

Note that our sample size is relatively small for applying structural equation modeling. We have, in total, 13 indicators. According to Nunnally and Bernstein (1994), the rule of thumb requires 10 observations per indicator to set a lower bound for the adequacy of sample sizes (see Marcoulides & Saunders, 2006, for a discussion about the problems with heuristics and related issues). We recognize that our sample size is only at the lower end of satisfactory zone. To quantify the potential impact of the sample size on the result, we computed the power of hypothesis test of close-fit based on RMSEA (MacCallum, Browne, & Sugawara, 1996). In particular, we were interested in the likelihood of rejecting the null hypothesis RMSEA ≤ 0.05 if the true value of RMSEA was actually 0.08 or 0.10. Given the significance level α = 0.05, the probabilities we found are respectively 77 percent for alternative hypothesis RMSEA = 0.08 and 99 percent for RMSEA = 0.10. The powers are close to or far above the desired level of 80 percent for typical hypothesis testing, which further indicates that the satisfaction of our sample size is reasonable.

We hypothesized six alternative measurement models based on the substantive nature of the correl ations among the first-order constructs. Among them, the group-factor model (Model A in Figure 3) stipulates that PS, PP, PI, and PA account for the variance of all 17 scale items, and that every pair of constructs correlates (although the correlation is not strong enough to justify a merger). The sec ond-order model (Model B in Figure 3), on the other hand, stipulates that PS, PP, PI, and PA accou nt for the variance of the 17 items and the second-order factor – trust – accounts for the covariance among PS, PP, PI, and PA.

Before testing the hypothesized models, we tested each first-order factor, first in isolation and then in pairs. We removed the items PA4, PA5, PS4, and PP3 (see Appendix A-1) in the process due to their low loadings to the respective factors. Then, using LISREL 8, we tested the hypothesized models based on the remaining 13 items. We found that both the first- and second-order models were strongly supported, whereas other alternatives were rejected based on the most frequently cited model-fit indices. Appendix B-1 shows the structures and estimated parameters of the supported models , Table 1 shows the model fit indices.

Table 1. Model Fit Indices

	Group-factor model	Second-order model	Threshold
χ^2(df)	57.24 (59)	63.13 (61)	
χ^2 Sig.	0.541	0.401	> 0.05
χ^2/df	0.970	1.035	< 2.00
NFI	0.935	0.927	> 0.90
CFI	0.998	0.992	
GFI	0.951	0.947	> 0.90
AGFI	0.925	0.920	> 0.90
NNFI	0.997	0.990	
RMR	0.041	0.046	< 0.05
RMSEA	0.000	0.014	< 0.05
P(RMSEA<.05)	0.977	0.956	> 0.90

For the measurement efficacies of the first-order model, we computed the composite reliability (Fornell and Larcker, 1981), Cronbach's alpha, AVE, and total variance explained for each of PS, PP, PI, and PA. All indices support the convergent validity of the first-order constructs with one exception: the AVE for PA was 0.45, which is a bit below the recommended 0.5 threshold (Fornell and Larcker, 1981). To test the discriminant validity of the constructs, we compared the first-order model with six alternative first-order models, each of which merges two of the first-order factors into a new one. We obtained significant χ^2 differences at α = 0.001 in all the comparisons and therefore declared strong support for distinctiveness of PS, PP, PI, and PA.

We then examined the efficacy of the second-order model in terms of target coefficient (Marsh & Hocevar, 1985), composite reliability, AVE, and observed total coefficient of determination. Here a target coefficient is the ratio of χ^2 (baseline model) to χ^2 (second order model) and composite reliability is the power of the second-order factor in explaining the variance of the first-order factors. We obtained a very high target coefficient value (0.91), which indicates that the introduction of the second-order factor into the baseline model does not significantly increase χ^2. The path loadings to the four first-order constructs were all above 0.70 (see Figure B in Appendix B-1), which demonstrates a strong convergent validity and reliability of the second-order measurement. The composite reliability and AVE were 0.89 and 0.66, respectively, which exhibits a good overall validity. The observed total coefficient of determination was 0.94, which indicates that a significant amount of the variance and covariance among PS, PP, PI, and PA is explained by the second-order construct and captured by the regression models. All of these indices suggest that the second-order factor model is acceptable as a better representation of the "true" factor structure.

4.2. Testing the Nomological Validity of Trust

Studies have shown that prior experience is the basis of trust (Blau, 1964; Hosmer, 1995; Kumar, 1996; Kumar, Scheer, & Steenkamp, 1995; Zucher, 1986). Trust is created or increased when experience is favorable, and ruined or reduced if it is negative. Accordingly, "prior experience" (PE) with the Internet is defined as a causal antecedent to trust. Using 5-point Likert scale ranging from "very negative" to "very positive", we asked each subject to describe their PE, and hypothesized that the more favorable the experience, the greater the trust.

On the other hand, trust is a prerequisite for behavior (Luhmann, 1979). It defuses concerns and thus encourages a willingness to take risks (Fukuyama, 1995). Many studies have empirically found that a consumer's trust in an Internet store impacts their willingness to patronize the store (Gefen, 2000; Jarvenpaa et al., 2000). Therefore, "purchase willingness" (PW) is defined as the extent to which a customer is willing to buy products or services from an Internet store, and it is measured by three scale items (see Appendix A-1). We hypothesized that a customer is more willing to buy from a store if they place a higher level of trust in the store.

To test the nomological validity of trust, we essentially made a comparison on what is a better mediator between PE and PW: the group-factor model or the second-order factor model? Thus, we embedded the two alternative models into a larger network of antecedent variable PE and consequent variable PW, which resulted in the two research models shown in Figure 7. Remember that these two models correspond to Model C and Model D respectively in Figure 3.

To perform the test using the efficiency-based approach, one must first ensure that all constructs are properly measured, including both antecedent and consequent variables. Since PE is observed directly, only PW's reliability whether it is a distinct concept from PS, PP, PI, and PA need to be checked. To this end, we computed Cronbach's alpha coefficient for the items of PW and conducted a factor analysis on 16 items: three items for PW and 13 items for PS, PP, PI, and PA. The alpha value was 0.79, which is above the acceptable threshold 0.7 (Nunnally, 1978), and which supports the convergent validity of PW. In the principle components analysis, we used the Kaiser eigenvalue criterion and extracted five factors that collectively explained 69.81 percent of the variance in all the items. We used Promax rotation and obtained the rotated factor matrix (see Appendix A-2). The results show that all of the items cleanly loaded onto the correct latent constructs, which therefore supports the factorial or discriminant validity of all the first-order variables.

Remember, as we discussed earlier on the difficulties with existing approaches, without freezing a measurement model, embedding antecedent and consequent variables will change the meaning of the constructs to be measured. For example, in Model B in Figure 7, if we do not freeze the measurement of trust, PW will become a new dimension of trust. After all, in a causal network, PW is at an identical position as the dimensions PP, PS, PI, and PA. The computer algorithm will treat PW as yet another dimension of trust for identification, which results in the change in the meaning of the concept we intend to measure.

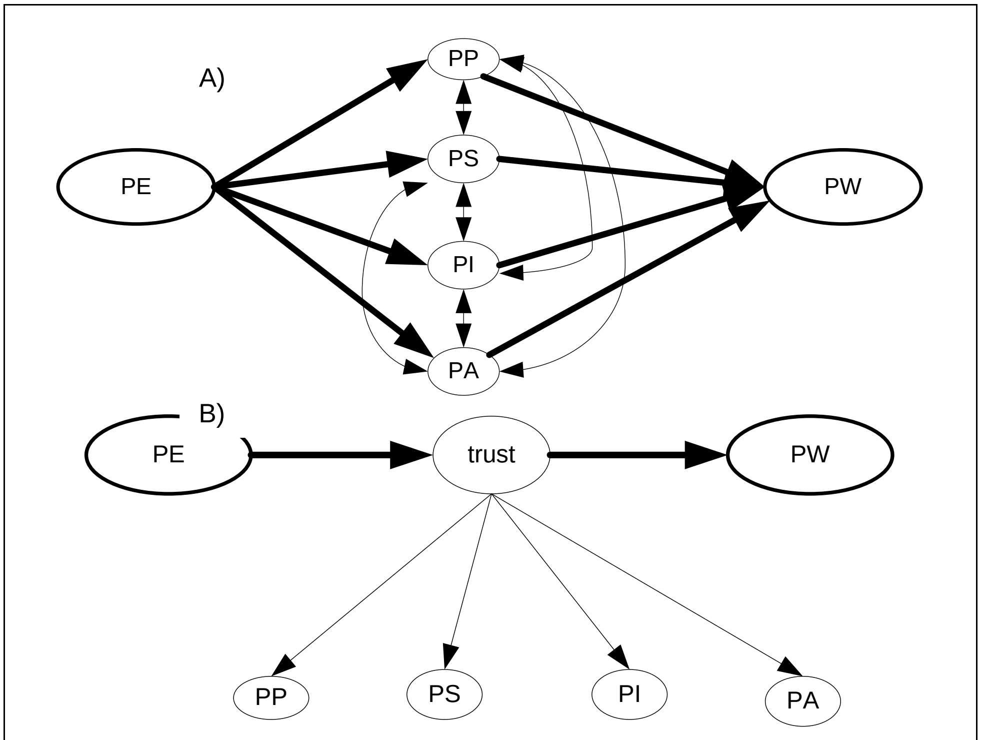

Figure 7. Testing the Nomological Validity of Trust: (a) Benchmark and (b) Target

To ensure that a measurement model does not change when embedded into a larger network of antecedent and consequent variables, we can fix all the free parameters in the established measurement model and test the larger network using LISREL. Alternatively, we can use factor scores obtained from the established measurement model and conduct ordinary linear regressions to do the same. For illustration, here we show multiple regressions. First, we obtained the factor scores for PP, PS, PI, PA, PW, and trust. We then treated them as proxy observable variables and conducted multiple linear regressions. PW is regressed against trust and compared with the regression of PW against PP, PS, PI, and PA. Figure 8 shows the regression results.

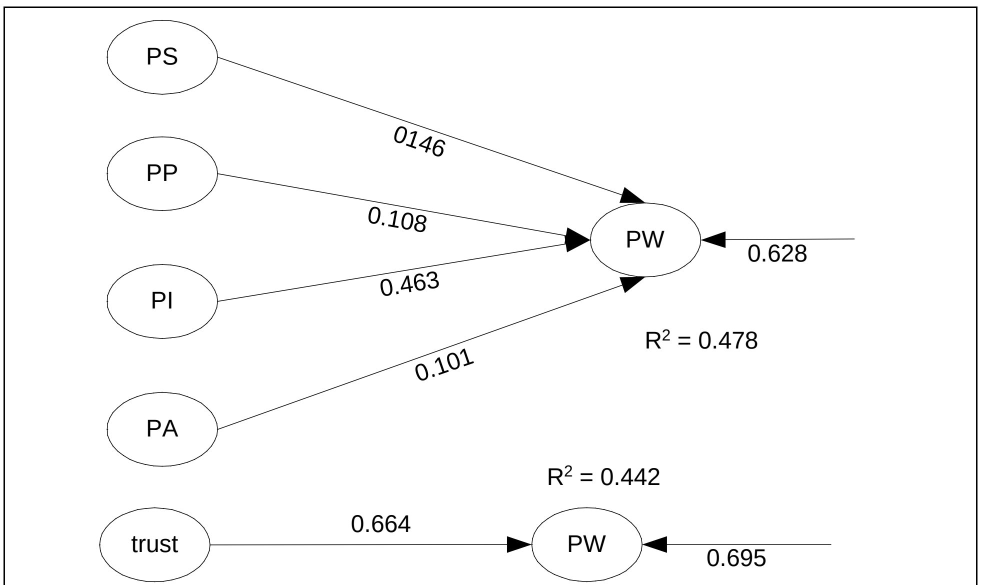

Figure 8. Computing the Predictive Efficiency of Trust

As indicated, the R^2 values for the two regressions were 0.442 and 0.478, respectively. This means that the first-order representation of trust – PS, PP, PI, and PA – acting as a joint predictor, explains 47.8 percent of the variance in PW, and the second order representation explains 44.2 percent. Therefore, the predictive efficiency of the second-order representation is the 0.442/0.478 = 92 percent. This implies that, among all of the variance in PW explained by PS, PP, PI, and PA, the second-order factor accounts for 92 percent of it.

We then conducted simple linear regressions using experience as the predictor and trust, PP, PI, PS, and PA, respectively, as the dependent variables. The results, shown in Figure 9, indicate that the R^2 value for the regression of trust against experience was 0.138. The AVE for the second-order factor was 0.66. Therefore, using the second-order factor as a mediator, the percentage of variance explained was $0.138 \times 0.66 = 0.091$. On the other hand, based on the loading and standard error estimates in Figure 9, we computed the AVE of the direct prediction of PP, PI, PS, and PA using the formula of Fornell and Larcker (1981) and obtained AVE (direct) = 0.100. Therefore, the mediating efficiency of the second-order trust was 91 percent. It implies that, among all the variance inherent in the trust concept that PE can explain, the same predictor PE can explain 91 percent when we use the second-order factor to represent the concept.

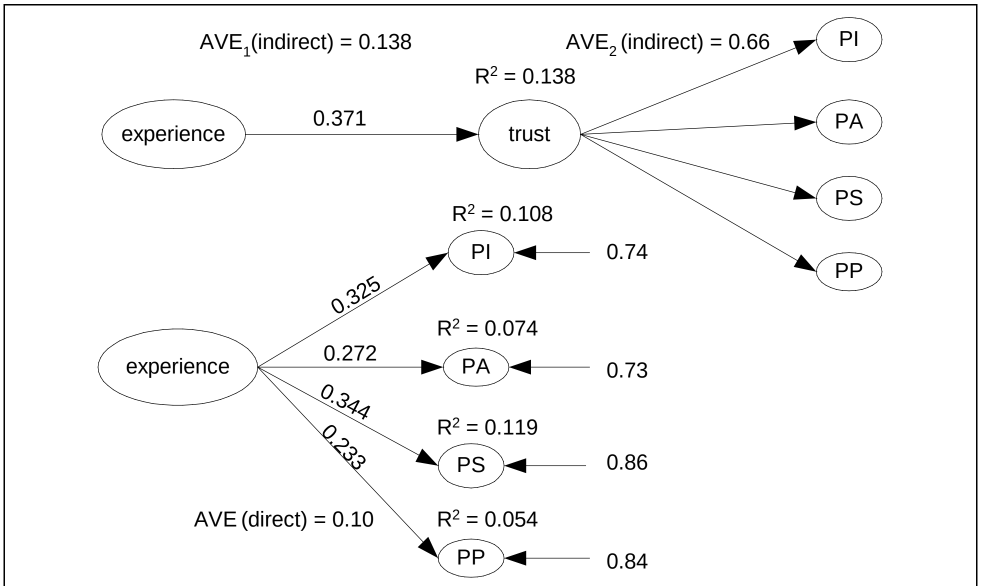

Figure 9. Computing the Mediating Efficiency of Trust

5. Conclusion

Higher-order constructs are widely-used in MIS research. Testing their nomological validity is to ensure that we measure the concepts that we intend to measure rather than artificial entities that are generated by statistics. Yet little research has been devoted to this topic and existing approaches all have difficulties and inconsistencies. To our knowledge, this paper is the first one in the literature providing a thorough treatment on the topic and a new systematic test approach. It also applies the approach to the concept of trust in an online store, one of most complex phenomena in IS.

Our approach is built on the premise that, if a higher-order construct captures the meaning of an intended concept, it should be as good as its first-order dimensions in explaining or predicting other related phenomena, and it should also be nearly equivalent to its first-order dimensions in representing a concept to be explained or predicted by others. Our test includes notions of predictive efficiency and mediating efficiency. Predictive efficiency measures how well a higher-order factor relative to all the first-order factors explains the variance in a consequent variable. Mediating efficiency measures how well a causal antecedent explains the total variance inherent in a concept indirectly through a higher-order factor as a mediator as compared to when it does so directly without the mediator.

To illustrate our approach, we propose a new measurement model of trust. We conceptualize trust as a second-order factor, which is reflected by four first-order dimensions: perceived security, perceived privacy, perceived inopportunism, and perceived accuracy. We then embed trust into a nomological network of prior experience as a causal antecedent, and purchase willingness as a consequent. Using the data we collected from 173 subjects, we validated the measurement model using usual reliability and validity tests. We then found that the predictive efficiency of the second-order trust is 92 percent and the mediating efficiency is 91 percent. The results suggest that the second-order abstraction can explain 92 percent of the variance in purchase willingness that can be explained by the first-order factors. Prior experience can explain 91 percent of the variance in trust when it is modeled as a second-order factor relative to when it is modeled as first-order factors. Therefore, we found strong support for the nomological validity of trust as a second-order factor.

Before discussing the contributions and implications of this study, we should note its limitations. First, our sample size is at lower end of the satisfactory zone, and one statistical power, the probability of rejecting null hypothesis of RMSEA $\leq$ 0.05 when the true value was 0.08, was relatively low at 77 percent. Second, the use of student participants may affect the external validity of the study (Gordon, Slade, & Schmitt, 1986). The same concern also affects other similar studies (Gefen, 2000; Jarvenpaa et al., 2000; McKnight et al., 2002). To overcome this limit, we carefully simulated and controlled many parameters to enhance the odds that our sample is representative of online consumers. A related concern regarding the use of students is that the participants did not truly make themselves vulnerable by risking their own money on a purchase. For this reason, we measured trust beliefs rather than trust-related behavior. Nevertheless, a further replication using online customers making actual purchases might be worthwhile. Given that measurement models could vary across samples and research contexts (Doll & Xia, 1997), such an additional test would have value in its own right.

A third limitation is related to the determination of the threshold for predictive and mediating efficiencies. This study established that both predictive and mediating efficiencies have an upper bound of 100 percent. Subsequent questions may arise regarding what are acceptable threshold values and what a threshold value means to a researcher. Other related questions also need to be addressed. For example, if a researcher performs this analysis and finds that their levels of efficiency are at 65 percent or some other low numbers that may indicate concern, how would they be able to improve the efficacies?

We tentatively suggest a threshold to be above 90 percent for predictive efficacy and 80 percent for mediating efficacy, but acknowledge that that more research, possibly involving computer simulations, is needed to recommend a more convincing or objective specification. Based on the limited data for validating our second-order measurement of trust, we conducted a few simulations by assuming that trust is made of different component dimensions. The result in Appendix C shows high predictive efficacies (> 90 percent) and mediating efficacies (> 80 percent) when the dimensions are sampled from factors PS, PI, PA, and PP, which are truly reflective of trust. For example, assuming trust is made of PS, PA, and PP, the predicative efficacy reaches 99 percent. However, by including Internet experience (NE) as a dimension of trust, which we know it is not, the efficacy values go below 90 percent. For example, assuming trust is made of one PI and NE, the predictive efficacy becomes 72 percent, and mediating efficacy reaches as low as 70 percent.

Note that nomological validity tests whether the construct captures the intended concept as reflected by its dimensions. It does not tell how well the construct is measured. Therefore, by removing some dimensions from a construct, we may end up with a poor measurement model but the construct can still capture the intended concept well. Thus, we see high efficacies even though we removed a dimension from the original model. In fact, we may see even higher efficacies because the measured construct becomes more coherent relative to the one that taps more diverse aspects. On the other hand, by adding a dimension unrelated to the construct, it is nature that the meaning of the concept changes and poor nomological validity results.

In theory, both mediating and predictive efficacies are between 0 and 100 percent. In practice, finding or generating samples that have low efficacy indices is a challenge partially due to the technical limitation of sampling multivariate normal distributions; existing algorithms impose positive definite covariance matrices as limitation, while our simulation design often involves degenerate normal distributions that do not have positive definite covariance matrices. Within this limitation, we have found that predictive efficacies can be as low as 50 percent so far. We are currently developing a simulation design for the specification of probability-based threshold and the generation of lower efficacy values.

A fourth limitation is the applicable scope of our method. Predictive and mediating efficiencies are proposed for reflective higher-order models, including those whose first-order factors may be formative. However, questions remain about whether they can be applied to formative higher-order constructs. A formative construct is considered a mixed bag of independent first-order concepts. Often, researchers who use formative constructs are interested in these first-order constructs instead of the higher-order construct and their nomological network of other variables. Thus, the measurement of the higher-order factor is not an issue. Nevertheless, in cases when the measurement of higher-order formative constructs is important, nomological validity may still be necessary, and it is up to future research to find such tests.

A fifth limitation is with the validation of the measurement model of trust. Existing models in the literature borrowed formative factors such as competence, benevolence, and integrity from Mayer et al. (1995) and use them incorrectly as reflective dimensions. We developed our model from the ground up in order to obtain a reflective structure. Some dimensions may be undesirable or debatable. For example, the notion of perceived reliability came up but is not included as a dimension. We originally considered it in the place of perceived accuracy. However, we realized that it took at least two or more interactions for the participant to meaningfully respond to the questions regarding reliability. Since our participants had only one instance of doing a fake purchase online, we settled with perceived accuracy.

This study has made a few contributions to the MIS literature. First, our approach to nomological tests greatly improves ad hoc ones in the literature. Like the nomological test of first-order constructs by Salisbury et al. (2002), our test emphasizes freezing a measurement model or the meaning of a concept in nomological tests. However, our two-step execution makes the idea applicable to higher-order models. Like Steward and Segars (2002), our test uses an extended group-factor model as the benchmark. However, our test has consistent and generalizable criteria and also ensures that the meaning of a concept does not change before and after it is embedded into a larger nomological network. Unlike McKnight et al. (2002), our test does not emphasize the absolute correlation between a higher-order construct and a related concept; such a correlation guarantees only that the higher-order construct is related but does not guarantee that it is not an artificial entity.

Second, we applied our new nomological approach to an empirically established second-order measurement model of trust in Internet stores. As demonstrated by our analysis, this model of trust captures the actual meaning of trust rather than that of an artificial entity. Our measurement of trust is the first in the literature that is developed from customers' real concerns about online transactions. It is consistent with the standard second-order reflective structure. Existing models of trust (Benbasat & Wang, 2005; McKnight & Chervany, 2002) defined trust as beliefs on a merchant's or agent's competence, benevolence, and integrity. Besides the issue of whether online merchants can be characterized as benevolent or not, these beliefs do not form the standard reflective structure of higher-order models according to the criteria of MacKenzie et al. (2005), and were originally formative in the integrative model of Mayer et al. (1995).

Most importantly, this study provides us with a better understanding of the concept of nomological validity – a concept that is applicable not only to trust but also other higher-order measurement models. It has significant implications to theory testing in IS research and practically measuring IS constructs such as system usage (Straub, Limayem, & Karahanna, 1995) and end-user satisfaction (Doll & Torkzadeh, 1991); establishing a substantive validity of a construct before testing its construct validity may lead to the accumulation of knowledge that later must be discarded (MacKenzie et al., 2005).

Acknowledgements

The authors would like to thank Senior Editor Carol Saunders and the three anonymous referees for their constructive comments and suggestions. The first author would also like to thank Dr. Steven Karau for his help in developing initial instruments.

References

Anderson, R. D., & Rubin, H. (1956). Statistical inference in factor analysis. *Proceedings of the Third Berkeley Symposium of Mathematical Statistics and Probability*, *5*, 111-150.
Bagozzi, R. P., & Edwards, J. R. (1998). A General approach for representing constructs in organizational research. *Organizational Research Methods*, *1*(2), 45-87.
Barber, B. (1983). *The logic and limits of trust.* New Brunswick, NJ: Rutgers University Press.
Benbasat, I., & Wang, W. (2005). Trust in and adoption of online recommendation agents. *Journal of the Association for Information Systems*, *6*(3), 72-101.
Blau, P. (1964). *Exchange and power in social life*. New York, NY: Wiley.
Bohrnstedt, G. W. (1970). Reliability and validity assessment in attitude measurement. In G. F. Summers (Ed.), *Attitude measurement* (pp. 80-99). Chicago, IL: Rand McNally.
Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, *110*(2), 305-324.
Butler, J. K. J. (1991). Toward understanding and measuring conditions of trust: evolution of conditions of trust inventory. *Journal of Management*, *17*(3), 643-663.
Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81-105.
Cattell, R. B. (1952). *Factor analysis: An introduction and manual for the pschologist and social scientist*. New York: Harper.
Chin, W. W. (1998). Issues and opinion on structural equation modeling. *MIS Quarterly*, *22*(1), vii-xvi.
Churchill, G. A., Jr. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, *16*(1), 64-73.
DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research*, *3*(1), 60-95.
DeVellis, R. F. (2003). *Scale development: Theory and applications*. Thousand Oaks: Sage Publications.
Doll, W. J., & Torkzadeh, G. (1991). The measurement of end-user computing satisfaction: Theoretical and methodological issues. *MIS Quarterly*, *15*(1), 5-10.
Doll, W. J., & Xia, W. (1997). A confirmatory factor analysis of the end-user computing satisfaction instrument: A replication. *Journal of End-User Computing*, *9*(2), 24-31.
Doney, P. M., & Cannon, J. P. (1997). An examination of the nature of trust in buyer-seller relationships. *Journal of Marketing*, *61*(2), 35-51.
Edwards, J. R. (2001). Multidimensional constructs in organizational behavior research : An integrative analytical framework. *Organizational Research Methods*, *4*(2), 144-192.
Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*(2), 155-174.
Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobserved variables and measurement error. *Journal of Marketing Research*, *18,* 39-50.
Fukuyama, F. (1995). *Trust: The social virtues & the creation of prosperity*. New York, NY: The Free Press.
Gefen, D. (2000). E-commerce: The role of familiarity and trust. *Omega: International Journal of Management Sciences*, *28*(6), 725-737.
Gordon, M. E., Slade, L. A., & Schmitt, N. (1986). The 'science of the sophomore' revisited: From conjecture to empiricism. *Academy of Management Review*, *11*(1), 191-207.
Hoffman, D. L., Novak, T. P., & Peralta, M. (1999). Building consumer trust online. *Communications of the ACM*, *42*(4), 80-85.
Hosmer, L. T. (1995). Trust: The connecting link between organization theory andphilosophical ethics. *Academy of Management Review*, *20*(2), 379-403.
Jarvenpaa, S. L.,Tractinsky, N., & Vitale, M. (2000). Consumer trust in an Internet store. *Information Technology and Management*, *1*(1), 45-71.
Jiang, J. J., Klein, G., & Carr, C. L. (2002). Measuring information system service quality: SERVQUAL from the other side. *MIS Quarterly*, *26*(2), 145-160.
Kappelman, L. A. (1995). Measuring user involvement: a diffusion of innovation perspective. The *Database for Advances in Information Systems*, *26*(2/3), 65-86.
Kumar, N. (1996). The power of trust in manufacturer-retailer relationships. *Harvard Business Review*, *74*(6), 92-106.

Kumar, N., Scheer, L. K., & Steenkamp, J. E. M. (1995). The effects of supplier fairness on vulnerable resellers. *Journal of Marketing Research*, *32*(1), 54-65.
Lee, M. K. O., & Turban, E. (2001). A trust model for consumer Internet shopping. *International Journal of Electronic Commerce*, *6*(1), 75-91.
Lewis, J. D., & Weigert, A. (1985). Trust as a social reality. *Social Forces*, *63*(4), 967-985.
Luhmann, N. (1979). *Trust and power.* Great Britain: John Wiley and Sons.
MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130-149.
MacKenzie, S. B., Podsakoff, P. M., & Jarvis, C. B. (2005). The Problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *Journal of Applied Psychology*, *90*(4), 710-730.
Marcoulides, G. A., & Saunders, C. (2006). PLS: A silver bullet? *MIS Quarterly*, *30*(2), iii-ix.
Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First and higher order factor models and their invariance across groups. *Psychological Bulletin*, *97*(3), 562-582.
Mayer, R. C., Davis, J. H., & Schoorman, D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709-734.
McKnight, D. H., & Chervany, N. L. (2002). What trust means in e-commerce customer relationships: An interdisciplinary conceptual typology. *International Journal of Electronic Commerce*, *6*(2), 35-60.
McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, *13*(3), 334-359.
Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. Thousand Oaks: Sage Publications.
Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.
Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research*, *23*, 51-67.
Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, *23*(3), 393-404.
Salisbury, W. D., Chin, W. W., Gopal, A., & Newsted, P. R. (2002). Research report: Better theory through measurement – developing a scale to capture consensus on appropriation. *Information Systems Research*, *13*(1), 91-103.
Segars, A. H., & Grover, V. (1998). Strategic information systems planning success: An investigation of the construct and its measurement. *MIS Quarterly*, *22*(2),139-163.
Smith, J. B., & Barclay, D. W. (1997). The effects of organizational differences and trust on the effectiveness of selling partner relationships. *Journal of Marketing*, *61*, 3-21.
Son, J.-Y., & Kim, S. S. (2008). Internet users' information privacy-protective responses: A taxonomy and a nomological model. *MIS Quarterly*, *32*(3), 503-529.
Stewart, K. A., & Segars, A. H. (2002). An empirical examination of the concern for information privacy instrument. *Information Systems Research*, *13*(1), 36-49.
Straub, D. W., Limayem, M., & Karahanna, E. (1995). Measuring system usage: Implications for IS theory testing. *Management Science*, *41*(8), 1328-1342.
Taylor, S., & Todd, P. A. (1995). Understanding information technology usage: A test of competing models. *Information Systems Research*, *6*(2), 144-176.
Thompson, R. L., Higgins, C. A., & Howell, J. M. (1991). Personal computing: Toward a conceptual model of utilization. *MIS Quarterly*, *15*(1), 125-142.
Thurstone, L. L. (1935). *The vectors of mind.* Chicago: University of Chicago Press.
Tygar, J. D. (1998). Atomicity in electronic commerce. *Mixed Media*, 32-43.
Wang, H. Q., Lee, M. K. O., & Wang, C. (1998). Consumer privacy concerns about Internet marketing. *Communications of the ACM*, *41*(3), 63-70.
Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, *12*(4), 5-28.
Zucher, L. G. (1986). Production of trust: Institutional sources of economics structure, 1840-1920. *Research of Organizational Behavior*, *8*, 53-111.

Appendices

Appendix A.

Table A-1. Initial Measurement Items for Trust

Perceived security	
PS1	I believe that shopping on this Internet store is just as safe as placing an order by phone
PS2	It is just as safe to make a credit card purchase at this Internet store as it is to make one in person
PS3	The data transmission between my computer and this Internet store is safe
PS4	This Internet store is capable of preventing illegal access
Perceived privacy	
PP1	I trust that this Internet store will keep customer information confidential
PP2	I am afraid that this internet store might misuse my personal information (Reverse)
PP3	I think it’s likely that this store would sell my personal data to others (Reverse)
PP4	This store can be trusted to keep the identities of its customers private
Perceived inopportunism	
PI1	I have confidence in this store
PI2	I trust this store to keep my best interests in mind
PI3	I trust this store to act professionally
PI4	Something about this store strikes me as deceptive and misleading (reverse)
Perceived accuracy	
PA1	I would trust this store to deliver exactly what I order
PA2	The Internet store can be trusted to fulfill my order accurately
PA3	The store will not overcharge my credit/debit account
PA4	I believe that the online product and service information is accurate
PA5	This store will make corrections if my order is in error
Purchase willingness	
PW1	I am willing to make a purchase from this store
PW2	I would like to come back to this store next time when I need to buy something
PW3	I do not want to buy anything from this site in the future

Note: All items are in 5-point scale anchored by “strongly disagree” and “strongly agree”.

Table A-2. Factor Analysis

	Factors				
	PI	**PP**	**PW**	**PA**	**PS**
PI1	**.53**	.07	.33	.09	-.05
PI2	**.69**	.14	.15	-.15	.02
PI3	**.82**	.01	-.01	.10	.05
PI4	**.74**	.02	.00	.09	.00
PA1	.14	.02	-.07	**.80**	-.14
PA2	.13	.06	-.06	**.72**	-.01
PA3	-.11	-.12	.11	**.83**	.06
PS1	.29	-.14	-.17	-.13	**.91**
PS2	-.18	.16	.12	.06	**.82**
PS3	-.09	.07	.03	.34	**.59**
PP1	-.14	**.83**	.04	.03	.13
PP3	.16	**.82**	-.03	-.06	-.14
PP4	.18	**.76**	-.05	.00	.08
PW1	.05	.00	**.70**	.00	.12
PW2	-.11	.11	**.96**	-.01	-.17
PW3	.28	-.22	**.73**	-.05	.10

Principal Component Analysis: Promax with Kaiser Normalization

Appendix B.

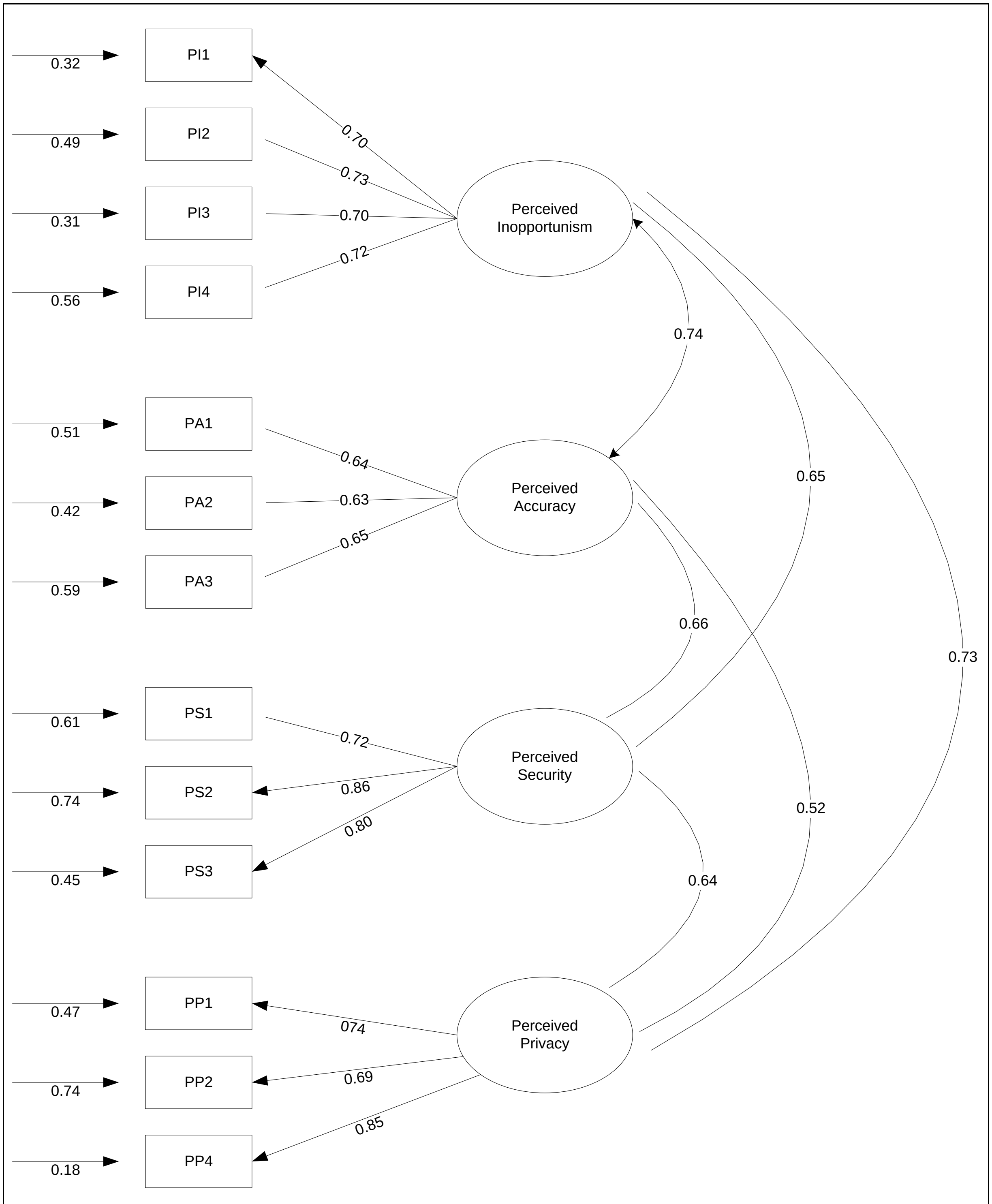

Figure B-1. A Four First-Order Factor Model of Trust (χ^2 / df = 0.970, P(RMSEA < .05) = 0.977)

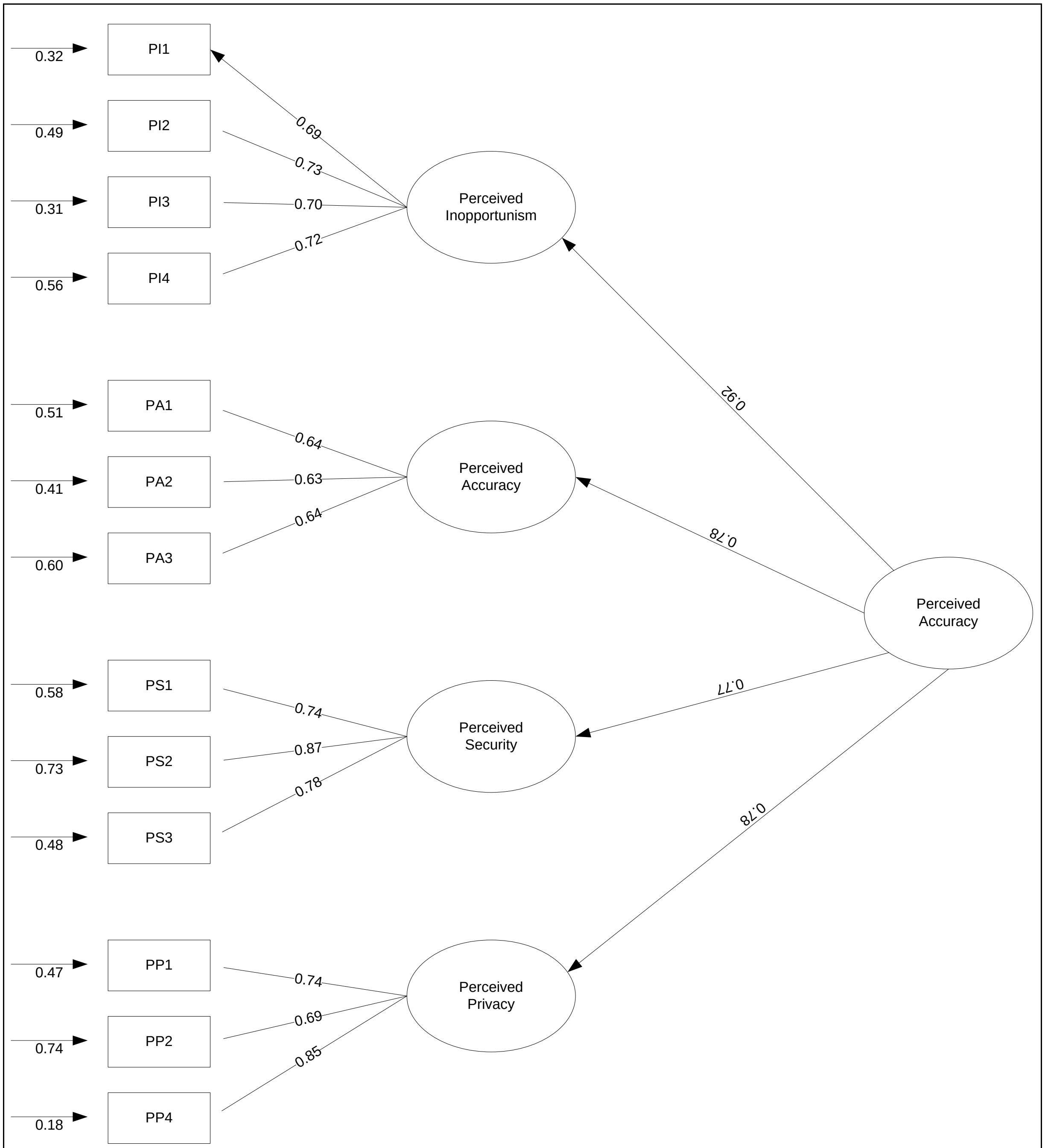

Figure B-2. A Second-Order Factor Model of Trust (χ^2 / df = 1.035, P(RMSEA < .05) = 0.956)

Appendix C. Simulations with Different Compositions of Trust

Table C-1. Impacts on Predictive Efficacy

Trust composition	R^2 (base)	R^2 (target)	Predictive efficacy
PS, PI, PA, PP	0.478	0.442	92%
PI, PA, PP	0.464	0.433	93%
PS, PA, PP	0.373	0.371	99%
PS, PI, PA	0.471	0.427	91%
PS, PI, PP	0.471	0.428	91%
PS, PI, PA, PP, NE	0.478	0.422	88%
PS, NE	0.245	0.208	85%
PI, NE	0.437	0.317	72%
PP, NE	0.248	0.198	80%
PA, NE	0.222	0.167	75%

Table C-2. Impacts on Mediating Efficacy

Trust composition	AVE(direct)	AVE_1(indirect)	AVE_2(indirect)	Mediating efficacy
PS, PI, PA, PP	0.100	0.138	0.660	91%
PI, PA, PP	0.116	0.112	0.767	74%
PS, PA, PP	0.111	0.128	0.710	82%
PS, PI, PA	0.140	0.147	0.756	79%
PS, PI, PP	0.123	0.131	0.754	81%
PS, PI, PA, PP, NE	0.128	0.164	0.645	82%
PS, NE	0.148	0.171	0.705	81%
PI, NE	0.164	0.154	0.743	70%
PP, NE	0.107	0.102	0.738	70%
PA, NE	0.141	0.119	0.770	65%

About the Authors

Liping LIU is Professor of Management and Information Systems at the University of Akron. He received his Bachelor of Science in Applied Mathematics in 1986 from Huazhong University of Science and Technology, Bachelor of Engineering in River Dynamics in 1987 from Wuhan University, Master of Engineering in Systems Engineering in 1991 from Huazhong University of Science and Technology, and Ph.D. in Business in 1995 from the University of Kansas. His research interests are in the areas of Uncertainty Reasoning and Decision Making in Artificial Intelligence, Electronic Business, Systems Analysis and Design, Technology Adoption, and Data Quality. His articles have appeared in *Decision Support Systems*, *European Journal of Operational Research*, *IEEE Transactions on System, Man, and Cybernetics*, *International Journal of Approximate Reasoning, Information and Manageme*nt, *Journal of Optimization Theory and Applications*, *Journal of Risk and Uncertainty*, and others.

Chan LI is an assistant professor in accounting at University of Pittsburgh. She has published papers in wide range of accounting and information system journals, such as *Journal of Accounting and Economics*, *Contemporary Accounting Research*, *Auditing: A Journal of Practice & Theory*, and *MIS Quarterly*.

Dan ZHU is Associate Professor of Information Systems and Computer Science at Iowa State University. She received her PhD from Carnegie Mellon University. Her research emphasizes the development of computational and analytical models to support business decision making. Dr. Zhu has published papers in *Proceedings of National Academy of Sciences, ACM Transactions, INFORMS Journal on Computing, Information System Research, Decision Sciences, Annals of Statistics, Naval Research Logistics, Annals of Operations Research, Decision Support Systems*, among others.